UNITED
NATIONS

E

Economic and Social Council

Distr.
LIMITED

E/CONF.83/L.25
21 January 1991

ENGLISH ONLY

TWELFTH UNITED NATIONS REGIONAL
CARTOGRAPHIC CONFERENCE FOR
ASIA AND THE PACIFIC
Bangkok, 20-28 February 1991
Item 7 (b) of the provisional agenda*

# CARTOGRAPHIC DATA RETRIEVAL, ANALYSIS, DEPICTION, PRESENTATION AND PRODUCT GENERATION: DIGITAL DATA BASES

## ARC digitized raster graphics and their application

Paper submitted by the United States of America**

* E/CONF.83/1.

** Prepared by Donna H. Foley and Jere R. Cadozet, Defense Mapping Agency.



READEX FICHE

ARC Digitized Raster Graphics (ADRG) are digitized replicas of hard copy source maps and charts that the Defense Mapping Agency produces on Compact Disc-Read Only Material (CD-ROM) for distribution.

The ADRG process converts the hard copy source datum and projection to the World Geodetic System 1984 (WGS-84) and the Equal ARC-Second Raster Chart/Map (ARC) System which permits a worldwide seamless data base for map data of a given scale.

Products currently included are: Operational Navigation Charts (ONCs) at 1:1,000,000, Tactical Pilotage Charts at 1:500,000, Joint Operations Graphics (JOGs) at 1:250,000, Topographic Line Maps (TLMs) at 1:50,000 and Jet Navigation Charts (JNCs) at 1:2,000,000. Many CD-ROMs of small scale maps are available for public scale.

ADRG data is best suited for use as raster background images for geographic information system applications. Within U.S. Department of Defense, ADRG currently supports the Navy's AV8B program and the Army's Maneuver Control System.

ADRG Product Description

Definition

ADRG are digital images derived from printed maps and charts and are distributed on CD-ROM. ADRG is comprised of a raster image of a paper map or chart supported by ancillary raster images of chart-specific marginalia and text data such as datum, projection, and accuracy information.

ADRG raster map images are generated by transforming a scanned map from its original datum and projection to the WGS 84 on the projection defined in the Equal Arc system. Production of ADRG on WGS 84 in the ARC system permits generation of a worldwide seamless image database for maps of the same scale. Currently, ADRG coverage includes JNCs at 1:2,000,000, ONCs at 1:1000,000, TPCs at 1:500,000, and some JOGs at 1:250,000 and TLMs at 1:50,000.

ADRG data is organized into distribution rectangles which are groups of one or more same-scale maps. Distribution rectangle size is limited by the amount of data that will fit on a CD-ROM; the typical size does not exceed 600 megabytes. As a rule, ONCs, TPC, and JNCs are distributed one sheet per CD-ROM; six TLMs and four JOGs fit into a single distribution rectangle. On multi-sheet distribution rectangles, the joins between adjacent sheets are practically undetectable.

ADRG raster image data is collected as nominal 100 micron pixels in 24-bit red, green, and blue color. During the ADRG product design phase, 24-bit color was determined to be high enough resolution to withstand multiple resampling operations without significant image degradation.

For a given CD-ROM, top level data files include a text file describing the distribution rectangles that appear on the disc, as well as a color patch image file to use for color calibration on a user's display system. For a given distribution rectangle, there is a subdirectory containing text files of ancillary data that describe the ARC images, accuracy of the source map, and security/releasability information. The subdirectory also contains the ARC images themselves with the corresponding overview image.

The overview image, compressed from the actual 24-bit ARC image, is designed to fit in its entirety on a single screen display. It is the digital equivalent of a location diagram and provides a graphic index for rapid access to the full-resolution ARC image.

Each source map within a distribution rectangle also has its own complement of ancillary data files. These files include source-specific information such as map name, datum, essential aeronautical data, and map sheet boundary coordinates. Each

source map within a distribution rectangle also has a set of one or more raster images of map-specific marginalia. Examples include landmark feature symbols and boundary diagrams.

ADRG design reflects the Defense Mapping Agency's emphasis on standard exchange formats. All ADRG data and file directories are formatted in compliance with the ISO 8211 standard. International Standards Organization (ISO) 8211 is a self-describing file format flexible enough to accommodate a variety of data formats and is completely independent of the exchange medium. ISO 9660 has been selected as the volume and file structure for the ADRG CD-ROMs.

ARC System

The ARC system was designed for ease of computer display and for simple, direct computation of geographic coordinates from image pixel location.

As implemented in ADRG, ARC defines 16 non-polar bands called zones that lie between 80 degrees north and 80 degrees south latitude plus two additional zones at each pole (Figure 1). In the non-polar zones, ARC is based on an equirectangular projection (Synder, 1987, page 90); the spherical form of the azimuthal equidistant projection, polar aspect (Synder, 1987, page 191) forms the projection basis for the ARC system over the poles. The number of zones defined in ARC minimizes distortion in a given zone to no more than 18 percent maximum stretch or shrink. The datum used for ARC in ADRG is WGS 84).

Although the scale of any given application of the ARC system is dependent upon the scale of the map originally scanned, ARC has a simple relationship to angular distance for a given scale. Within a specified non-polar zone, any pixel represents the same angular distance in the east-west direction. Across zone boundaries in the north-south direction, each pixel represents the same angular distance in all non-polar zones.

ARC was also designed to support creation of seamless worldwide map image databases. As it is applied in ADRG, a 1024 pixel overlap between ARC zones allows programming of image display software to scroll seamlessly from zone to zone.

Pixels in ARC are organized into 128 x 128 pixel tiles. This tiling scheme, which is seamless across the worldwide ARC zones, permits development of optimized image data retrieval software that can access image data rapidly by tile.

ADRG Production Process

Production System

The ADRG production system consists of four 19" monitor color graphics data preparation workstations, one 44" x 60" large format

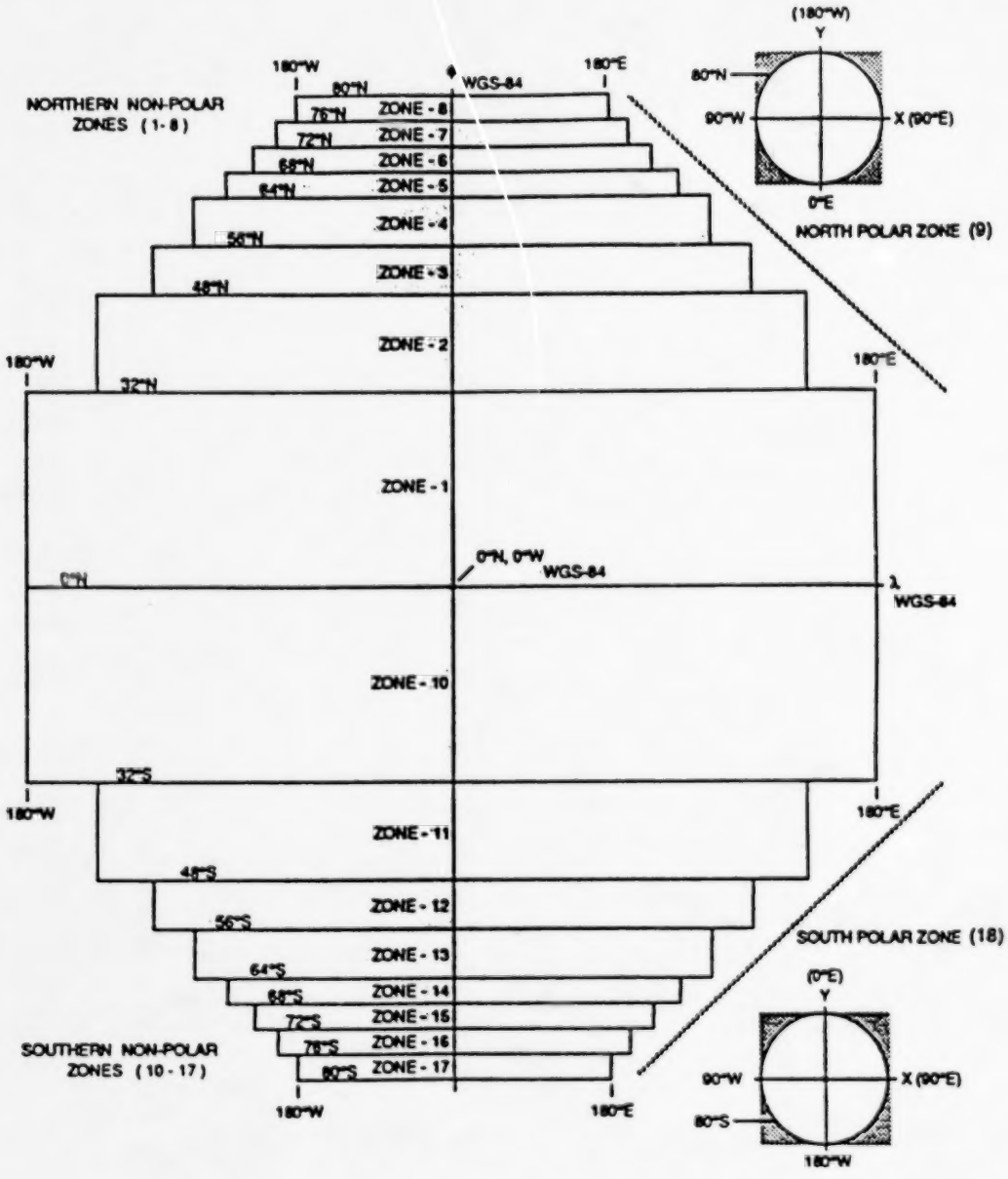


Figure 1. ARC System Zone Layout

drum scanner with interface processor, five transformation batch processors running at 10 million instructions per second with three gigabytes of disk space and 117 megabytes of memory, and two 27" monitor color graphics quality control workstations. All system components are connected to a dedicated network. Once input data is prepared for entry into the ADRG production system, data processing time for each CD-ROM ranges from five to seven hours, depending on the size and type of charts to be distributed on the particular CD-ROM.

Source Preparation

Although ADRG production is highly automated, initial chart preparation and data collection is manual. Cartographers manually update aeronautical charts with Chart Update Manual information to reflect conditions hazardous to navigation. Chart series in which adjacent charts overlap are also examined for any features or labels in an overlap region that need to be moved. Relevant chart parameters such as datum, projection, sheet corner coordinates, graticule spacing, and accuracies are collected from the chart itself and from other databases as text information. Other data collected identifies which marginalia such as boundary diagrams or elevations tint diagrams should also be captured as raster images. All text or ancillary data is entered into a microcomputer database from which a floppy disk is extracted for interface to the ADRG production system.

Processing

The ADRG production flow consists of seven main steps that are summarized in Figure 2. The production process begins with the entry of ancillary data into a data preparation workstation via floppy disk (Step 1). System resources are allocated during this step to ensure availability of adequate disk space for processing and to assist in automated tracking of data files through the production system. In addition, system operations may interactively verify and edit input ancillary data before data processing begins.

Upon completion of the data input process step, required data items are automatically validated and Metric Support Data (MSD) is generated. MSD is a set of polynomial coefficients for rapid transformation from WGS-84 or the ARC system to the source datum or projection for those ADRG users who require map information on its original datum and projection.

In the next process (Step 2), the paper map is mounted for scanning. The scanner collects swaths of 250 lines per inch around the circumference of the scanner drum. The 24-bit red, green, and blue data is deswathed and rotated so that the raster image is in row-major format. The result is a north-up, right-reading data file. During the scanning process, a copy of the

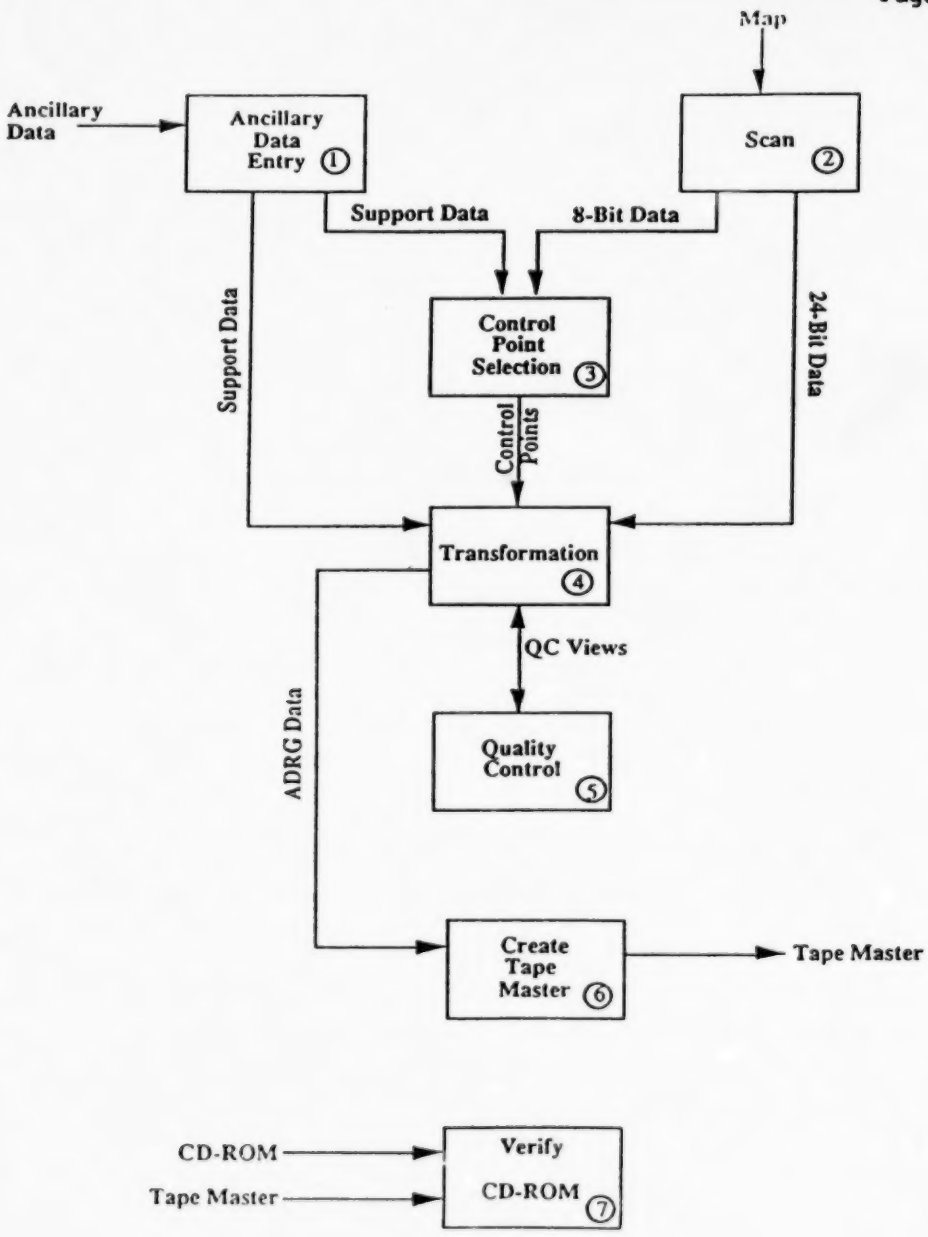


Figure 2. ADRG Production Flow

24-bit color data is condensed to 8-bit color data using a color table. The 24-bit data is sent over the network to a transformation node for later batch processing. The 8-bit data is sent to a data preparation node for control along with a condensed overview of the entire scanned area for use as a working file.

The system operator uses the 8-bit image data to select marginalia to be stored as raster images and to perform map registration (Step 3). Once the legends are selected, the control point selection software automatically generates a list of candidate areas for control using map corners and graticule spacing information defined in the ancillary data. Then the operator ties the geographic coordinates to the scanned image by visually identifying the (row column) coordinates that represent the four map corners. The software automatically presents to the operator an expanded view of the region surrounding each candidate area for sub-pixel refinement in identifying control points.

As each candidate point is measured, a sequential least squares adjustment is performed and the resulting residuals are displayed to aid the operator in detecting blunders. Once an adequate control solution is achieved, the control information and legend information is saved for the ARC transformation process step.

On the transformation nodes, control and legend information is merged with ancillary data, MSD, and the 24-bit color raster image data (Step 4). The ARC transformation algorithm is based on a single math model that dewarps the hardcopy map, performs the projection transformation to ARC and the datum transformation to WGS-84, and performs color resampling on the resultant pixels all in one step. In addition, the transformation software creates the overview image as a 16:1 spatial compression of the ARC image, modified by black enhancement algorithm to retain the map graticule.

The text and image files are placed in their proper file and directory structure according the ADRG U.S. Military Specification. The final step of the transformation process is formatting the ADRG files according to the ISO 8211 specification. A text data rep and an image tiling rep are then generated for quality control review.

Quality Control

ADRG quality control is performed on a large screen quality control workstation and consists of visual and quantitative checks (Step 5). The quality control cartographer visually reviews all marginalia images. The overview image is used to select a number of views from the full resolution ADRG image. Measurement of geographic graticule and grid intersections are made on the full resolution image. The software computes a measured point's displacement from its theoretically correct position, based on

known characteristics of the geographic graticule or grid. Ninety percent of the measured points must be within two pixels of their correct position.

Manufacturing and Distribution

Once an ADRG data set has passed quality control, an image of the final CD-ROM, including ISO 9660 formatting, is output on a set of nine-track master tapes (Step 6). The tapes are sent to a commercial compact disk manufacturer for mastering and replication. Sample copies of each ADRG CD-ROM title undergo a full or partial bit comparison to the original master tapes (Step 7). For further verification, an ADRG CD-ROM can be read back into a quality control workstation for visual display. After verification, ADRG CD-ROMs are sent to Defense Mapping Agency Combat Support Center, Philadelphia Distribution Center, for worldwide distribution.

## ADRG Applications

ADRG is designed to be a raster image database supported by cartographically significant text information that is adaptable to a variety of user-specific applications. The Defense Mapping Agency does not distribute any application software with ADRG CD-ROMs. Sufficient information is available in the U.S. Military Specifications for ADRG (also available through Defense Mapping Agency) to allow applications programmers to access any ADRG data appropriate to a specific user need.

ADRG raster map images are rapidly becoming the background and overview data for a variety of U.S. Department of Defense systems with geographic information systems characteristics. These U.S. Department of Defense systems primarily support mission planning activities, but some applications also support on-board and battle management systems.

One of the earliest ADRG applications was an on-board moving map display for the U.S. Navy AV8B (Harrier) and F/A-18. For this system, the U.S. Navy uses ADRG disks to derive a Compressed Aeronautical Chart which is then used to develop mission-specific flight path images. This ADRG-based product is put on the aircraft as an aid to the pilot. The AV8B and F/A-18 program currently use ADRG coverage of TPC.

At the battle management system level, U.S. Department of Defense systems use ADRG data to support activities over a specified area.

The U.S. Army's Maneuver Control System (MCS) is a mobile mission planning system for field command. MCS uses a compressed product derived from the ADRG images of existing maps.

The U.S. Navy's Tactical Air Mission Planning System can be based afloat on a land-based command center. ADRG data can be used as a background display for positioning aircraft or ships in a given area.

Joint Operational Tactical System is a command and control battle management system designed for a desktop computer shipboard or ashore. ADRG data will be used experimentally to support this interrogative tactical display system.

ADRG will also support the Navy's shipboard MH-53 mine-sweeping helicopter activity. Helicopters will report satellite Global Positioning System positions of mines to the carrier ship. The mine positions will then be displayed on an ADRG image as the data is collected. Although still in the prototype phase, ADRG coverage of Harbor and Approach and Coastal Charts is intended to support the planned MH-53 activity.

The widest use of ADRG data is for command and control mission planning over large areas.

The Joint Visually Integrated Display System is a fast interrogative tactical display that uses ADRG as a raster background. This system, located at the National Military Command Center, will be used by the Joint Chief of Staff to overlay intelligence information in vector format to support counternarcotics activities. Identical systems support the Drug Enforcement Agency and the Commanders in Chief for counter-narcotics.

The FULCRUM system used by the U.S. Air Force is another geographic information system-like system that supports intelligence and operations planning. FULCRUM can use ADRG data for background images and can exploit other digital mapping products such as Digital Terrain Elevation Data, Digital Feature Analysis Data, and Interim Terrain Data in conjunction with ADRG.

Other mission planning systems intending to use ADRG background raster data include Tactical Air Command's Mission Support System, which will use compressed ADRG data, and Strategic Air Command's Strategic Mission Data Preparation System.

## Conclusion

The ADRG production process represents a rapid means of incorporating existing hardcopy maps and charts into the realm of digital processing. As a data base, ADRG is well on the way to providing a wide range of raster background data; there are currently over 1,000 ADRG CD-ROMs available for distribution, many of which are available for public sale. Incorporation of ADRG data into any Geographic Information System is limited only by a user's application software.

Bibliography

1. Aerospace Warfare Systems, Defense Mapping Agency Systems Center, Aerospace Office, "Aerospace Programs Office Active Systems as of 21 September 1990.

2. Defense Mapping Agency Technical Report, Supplement to U.S. Department of Defense World Geodetic System 1984 Technical report; Part II - Parameters, Formulas, and Graphics for the Practical Application of WGS 84, DMATR 8350.2-8, 1 December 1987 (Second Printing).

3. Defense Mapping Agency, U.S. Military Specification ARC Digitized Raster Graphics (ADRG), MIL-A-89007, 22 February 1990.

4. International Standard 8211, Information Processing Specification for a Data Descriptive File for Information Interchange, 1985.

5. International Standard 9660, Information Processing - Volume and File Structure of CD-ROM for Information Interchange, First Edition, 15 April 1988.

6. Snyder, John P., Map Projections - A Working Manual, U.S. Geological Survey Professional Paper 1395, First Edition, 1987.

---

UNITED
NATIONS

E

Economic and Social Council

Distr.
LIMITED

E/CONF.83/L.26
21 January 1991

ENGLISH ONLY

TWELFTH UNITED NATIONS REGIONAL
CARTOGRAPHIC CONFERENCE FOR
ASIA AND THE PACIFIC
Bangkok, 20-28 February 1991
Item 7 (b) of the provisional agenda*

# CARTOGRAPHIC DATA RETRIEVAL, ANALYSIS, DEPICTION, PRESENTATION AND PRODUCT GENERATION: DIGITAL DATA BASES

## Digital cartographic data products and activities of the United States geological survey

Paper submitted by the United States of America**

* E/CONF.83/1.

** Prepared by Michael A. Domartz, United States Geological Survey.

The U.S. Geological Survey is the prime source of many kinds of digital cartographic data of the United States and its outlying areas. These data include (1) Digital Line Graph data, which contain information about boundaries, hydrography, hypsography, transportation, and other features found on the Geological Survey's topographic maps; (2) Digital Elevation Model data, which contain the elevations of a regular array of point observations; (3) land use and land cover data, which contain information about land use and cover patterns and other associated layers of information; and (4) geographic names data. The Geological Survey also coordinates the development of standards for digital cartographic data, conducts research in the areas of geographic information systems and related technologies, and distributes digital cartographic and other earth science data to users.

## Introduction

The U.S. Geological Survey is a source of many kinds of topographic and special-purpose maps of the United States and its outlying areas. It is also a source of digital map data.

Many public agencies, private organizations, and individuals need reliable cartographic and geographic knowledge about the Nation. To serve such needs, Geological Survey maps are compiled to exacting standards of accuracy and content.

Today, an ever greater extent of the Nation's landscape is being translated into digital map data. These data are collected from Geological Survey topographic maps or acquired from satellite or other non-photographic sensors. Such data help to create and update a comprehensive set of line, image, and thematic maps. The data can be analyzed to find answers to a host of cartographic and geographic questions.

## Digital Cartographic Data Products

Needs are growing rapidly for digital cartographic data to support mapmaking and geographic information systems. Moreover, sound Earth and water resources management now calls for computer-compatible cartographic and special-purpose data to provide a basis for analysis.

The Geological Survey produces digital cartographic data and distributes them through the National Digital Cartographic Data Base (NDCDB). The NDCDB includes:

- topographic line map data as Digital Line Graphs (DLG's),
- elevation data as Digital Elevation Models (DEM's), and
- land use and land cover data.

In addition, the Geographic Names Information System contains digital descriptions of places, features, and areas in the United States identified by proper names.

### Digital Line Graph (DLG) Data

DLG's are vector files of cartographic data made primarily by digitizing point locations, lines, and polygon outlines from map-separation materials. They are prepared from primary series (ranging in scales from 1:20,000 to 1:63,360) and from 1:100,000- and 1:2,000,000-scale source maps.

These data are a framework of reference for other data about the Earth and its natural resources. These data are useful for computer-assisted production of cartographic products. They are organized to support the analytical functions of geographic information systems.

The spatial data are topologically structured. Spatial relationships, such as adjacency and connectivity among data elements, are explicitly encoded.

In addition, DLG data elements may have coded attributes. These feature codes identify the major category, such as hydrography, to which a data element belongs and add other specific feature information, such as "shoreline" or "spring" or descriptions and parameters, such as "intermittent" or elevation of the water surface.

Work in progress includes the conversion of boundary, contour, hydrography, transportation, U.S. Public Land Survey System (PLSS), and other selected data to digital form.

DLG data derived from primary series topographic line maps are available for portions of all 50 States and Puerto Rico. For the 48 contiguous States and Hawaii, digitizing of boundaries is 25 percent complete; contours, 1 percent; hydrography, 6 percent; transportation, 5 percent; and PLSS, 26 percent. Coverage is available for approximately 2 percent of Alaska for each category. Most of the DLG data are distributed as 7.5-minute cells.

Transportation and hydrographic data were digitized from 1:100,000-scale quadrangle maps to meet the Bureau of the Census's needs for base map information for the 1990 Census. Coverage is available for all the 48 contiguous States and Hawaii. Boundary, PLSS, and contour data are being collected. These data are distributed in 30-minute blocks.

Boundary, hydrography, and transportation data were digitized from revised 1:2,000,000-scale sectional reference maps from the National Atlas of the United States of America. Coverage of the 50 States is available in 21 regional sections.

An improved data model, called Digital Line Graph-Enhanced (DLG-E), will soon be in use. DLG-E provides for the explicit representation of individual cartographic features, such as roads, counties, buildings, and streams, in addition to the topologically structured spatial data provided in the current DLG. This enhancement also provides a more extensive set of attributes and relationships for these features than exists in a DLG.

Digital Elevation Model (DEM) Data

DEM's are arrays of elevation data, usually at regular intervals. DEM's can be used to generate simulated views of a landscape; to calculate capacities of reservoirs and volumes of lava flows; to assess terrain in relation to other environmental factors; to predict brush and forest fire behavior; and to calculate slope for such applications as hydrologic process modeling to gauge the effect of precipitation on ground-water and surface-water runoff.

/...

The Geological Survey produces DEM's by computing elevations from digitized vector contour data, by scanning stereomodels produced from aerial photographs, or by resampling higher-resolution DEM's.

For DEM's corresponding in coverage to 7.5-minute topographic quadrangle maps, elevations are processed to produce data at 30-meter ground spacing with the Universal Transverse Mercator coordinate system as a framework. For Alaska, the DEM's corresponding in coverage to 15-minute topographic quadrangle maps are processed to produce data at 3-arc-seconds of longitude by 2-arc-seconds of latitude spacing. Coverage is available for 35 percent of the 48 contiguous States and Hawaii, and for approximately 8 percent of Alaska.

Recently, the Geological Survey began producing DEM's from 1:100,000-scale DLG contour data or from resampled 7.5-minute DEM's. These data are processed to produce data at 2-arc-second spacing, and are distributed in 30-minute blocks.

The Geological Survey also distributes DEM's produced from digital terrain elevation data collected by the Defense Mapping Agency from 1:250,000-scale maps. Elevations are interpolated every 3 arc-seconds of longitude and latitude, except in Alaska, where spacing varies. These DEM's are distributed in 1-degree cells. The 3-arc-second DEM's consist of more than 1.4 million elevation values. Coverage is available for all the 48 contiguous States and Hawaii, and most of Alaska and Puerto Rico.

Land Use and Land Cover Data

For the 48 contiguous States and Hawaii, land use and land cover data are made by digitizing land use polygon outlines from 1:250,000- and 1:100,000-scale source maps.

Work in progress includes digitizing land use and land cover data and associated map data, including hydrologic units, political units, and census county subdivisions. Federal and State land ownership has been digitized in selected areas.

These data are available in vector and raster forms. The vector data, encoded in the Geographic Information Retrieval and Analysis System (GIRAS) format, are topologically structured. The raster data, encoded in the Composite Theme Grid (CTG) format, are composed of an array of cells. Each cell is completely identified with attribute codes for all land use and land cover and associated map data.

Digital data are available for more than 80 percent of the 48 contiguous States and Hawaii. Completion is scheduled for 1992.

For Alaska, land cover data are generated using Landsat multispectral scanner digital data. These raster data are encoded in a format different from the CTG format described above. Approximate 2 percent of the coverage planned for Alaska is available.

Geographic Names Data

Information about places, features, and areas in the United States identified by proper names and geographic locations is distributed through the Geographic Names Information System. This national data base contains about 2 million entries; when completed between 1998 and 2002, it will hold about 5 million entries. The Geographic Names Information System is the Federal government's official source of geographic names as sanctioned by the U.S. Board on Geographic Names.

Coordination of Data Standards

The Geological Survey is also active in coordinating matters related to Federal digital cartographic data.

In 1980, the National Institute for Standards and Technology agreed to assign to the Geological Survey leadership in defining and maintaining Federal earth science standards.

Under this agreement, the Geological Survey is coordinating the development of a standard for the exchange of digital cartographic data, called the Spatial Data Transfer Standard. This Standard was developed through grants to the National Committee for Digital Cartographic Data Standards, which functioned under the auspices of the American Congress on Surveying and Mapping. The Federal Interagency Coordinating Committee on Digital Cartography also contributed to the development of the Standard.

The need to transfer spatial data between different systems is becoming increasingly important. The concern for common data formats and coding conventions cuts across all aspects of cartographic data handling. Users find it difficult and inefficient to use data because of differences in underlying concepts, terminology, and formats.

The proposed Spatial Data Transfer Standard provides a national standard for the exchange of digital spatial data. The standard includes definitions of terminology, a transfer specification, recommendations on reporting data quality, and terms and definitions of features commonly found on topographic maps and hydrographic charts. The Federal government, the private sector, and the academic community have been working for several years to develop this proposed standard.

Empirical testing of the Spatial Data Transfer Standard was completed in 1989. Improvements have been made to the Standard based on the test participants' suggestions. The Standard will be submitted to the National Institute for Standards and Technology in mid-1990 for designation as a Federal Information Processing Standard.

## Geographic Information System Facilities and Research

Spatial analysis is an emerging discipline being applied to geographic information systems at the Geological Survey. Over 500 Geological Survey scientists and engineers have received formal training in geographic information system technology. Geographic information system applications software is available on a nationwide network consisting of approximately 85 minicomputers, a mainframe computer, plus additional stand-alone microcomputer and workstation systems. Within the Geological Survey (and in more than 100 cooperative application and demonstration projects with other Federal and State agencies), geographic information system technology has been applied in fields as diverse as water quality, geologic mapping, minerals exploration, oil and gas development, earthquake and other natural hazards analysis, forest management, public health, environmental analysis, wildlife management, agriculture, soils mapping, urban planning, and the 1990 Census.

The Geological Survey operates 3 interdisciplinary geographic information system research laboratories located in Virginia, Colorado, and California to support cooperative applications development and to facilitate evaluation of advanced geographic information system technologies. Long-term research is designed to address land surface characterization for global change and other dynamic, time-dependent, process modeling activities. The Geological Survey's EROS Data Center in Sioux Falls, South Dakota, will serve as the national archive and data manager for land surface data for global change research.

## Earth Science Information and Sales

The Geological Survey's Earth Science Information Centers (ESIC's) offer nationwide information on earth science data holdings and sales service for Geological Survey map products and earth science publications. This network of ESIC's provides information about geologic, hydrologic, topographic, and land use maps, books, and reports; aerial, satellite, and radar images and related products; earth science and map data in digital form and related applications software; and geodetic data.

---

UNITED
NATIONS

E

Economic and Social Council

Distr.
LIMITED

E/CONF.83/L.27
21 January 1991

ENGLISH ONLY

TWELFTH UNITED NATIONS REGIONAL
CARTOGRAPHIC CONFERENCE FOR
ASIA AND THE PACIFIC
Bangkok, 20-28 February 1991
Item 7 (b) of the provisional agenda*

# CARTOGRAPHIC DATA RETRIEVAL, ANALYSIS, DEPICTION, PRESENTATION AND PRODUCT GENERATION: DIGITAL DATA BASES

## The digital chart of the world project

Paper submitted by the United States of America**

* E/CONF.83/1.

** Prepared by David M. Danko, Defense Mapping Agency.

91-01914 2116e (E)

/...

The Defense Mapping Agency is developing a new product, The Digital Chart of the World. It will provide world wide coverage using a topologically structured vector representation of the earth's land surface information on a micro-computer accessible storage medium. The ultimate objective of this project is to develop, refine and establish a suite of standards to support future Defense Mapping Agency digital data production and enhance the utility of digital spatial information in vector format. The Digital Chart of the World database will be used to test and define product standards; once the standards are complete the Digital Chart of the World will provide an example for future Defense Mapping Agency products.

The standards and the Digital Chart of the World product are being developed in cooperation with the United Kingdom, Canada, and Australia. International participants will review draft standards and specifications; exercise and evaluate prototypes; assist in the design of user pathways into the data; and develop hardware and software utilities to send, receive and exploit the data.

## Introduction

The Digital Chart of the World Research and Development project, initiated in 1988, has as its primary goal the development of Vector Product Standards for the provision of digital geographic information to U.S. Department of Defense and Intelligence Community activities.

As a secondary goal the project will develop the product specifications for a 1:1,000,000 vector map database. Over 275 Operational Navigation and Jet Navigation Charts will be digitized to provide a product with worldwide coverage. The new product, the Digital Chart of the World, will contain several gigabytes of topologically based vector data.

Environmental Systems Research Institute of Redlands, California is the prime contractor for the project. The project is a multinational effort. The development partners are the United States, Canada, Australia, and the United Kingdom. The military map producers from these four nations, the primary producers of the Operational Navigation Chart Series which will provide the information for the Digital Chart of the World, have joined in this effort to establish standards for data structures that will support direct use in geographic information systems, CRT displays, and special computer environments. The Digital Chart of the World will be the first product generated according to these standards. This paper discusses the process followed for accomplishing the project goals of Standards Development and Product Development.

## STANDARDS DEVELOPMENT

### Background

Increasing Demands for Digital Data Although the demands for Defense Mapping Agency's conventional paper products are as high as ever, the demand for video and digital data is increasing. There has been a significant increase in the number of automated systems supporting military functions. Some are captive or special purpose developments, some employ commercial off-the-shelf geographic information systems technology; all require machine readable mapping and charting data. Defense Mapping Agency currently produces video products and digital products which use a raster model (ARC Digitized Raster Graphics), but there is a need for "smart" vector data for selective display and geographic modeling and analysis. To support these functions, the data sets must carry topological relationships and be fully attributed.

Increasing Availability of Digital Data. The Defense Mapping Agency production process is undergoing a modernization program; an end to end Digital Production System will be phased in over the next 5 years. Foreign military mapping agencies are increasing their output of digital data; the nations participating in the Digital Chart of the World project are also extending their production systems to produce more digital data. The digital data pool is also being filled by scanning and digitization efforts in the private sector.

Demand for a "User Oriented" Structure. Presently there is a variety of structures and standards for digital vector data. Available non-commercial data transfer structures include:

TIGER - Topologically Integrated Geographic Encoding and Referencing files, Bureau of the Census.

DLG-E - Digital Line Graph Enhanced, US Geological Survey.

SDTS - Spatial Data Transfer Specification, Federal Geographic Data Committee.

DIGEST - Digital Geographic Information Exchange Standards, Digital Geographic Information Working Group, (an ad-hoc group of NATO nations).

MC&GFDES - Mapping, Charting, and Geodesy Feature Data Exchange Structure, Defense Mapping Agency's Digital Production System.

All of the referenced standards were designed to accommodate producer exchange of data, not user exchange. They support specific implementations rather than a general purpose; apply to bulk transfer of data, with built-in overhead and redundancy; have no requirement for communication between parties; and lack application software. In general, these standards do not support customer requirements for interactive use, tiling, indexing, and manipulation of the data.

With the projected increase in both the generation of digital data, and the number of applications for the data, the need for digital product generation and user exchange standards became evident, and motivated this research and development effort.

Approach

In October 1989 Environmental Systems Research Institute was contracted by Defense Mapping Agency to develop the Vector Product Standards and the Digital Chart of the World. The Vector Product Standards is designed to provide direct database utility and be

oriented to provide Defense Mapping Agency products for use in commercial geographic information systems.

Given enough time any group of engineers can come up with a standard. The quest here is to establish standards that can be readily embraced by the customer community. To accomplish this, two things are required: (1) well designed, standards that support a wide range of user demands and (2) a wealth of data produced according to the standards, to attract the users and promulgate the standards.

To meet this primary goal Environmental Systems Research Institute was required to develop a comprehensive suite of vector product standards. To insure that these standards would fulfill a broad range of needs, be as generic as possible, and work in a wide range of computing environments (both product and platform); the following design constraints were imposed:

(1) The standards must be designed to work within the constraints of a microcomputer (IBM 80286) with a limited amount of hard disk storage (30 MB). This would allow the use of vector products with inexpensive, and in many cases portable equipment. With the exception of tile size any design slated for PC class environment will also operate in workstation and larger environments with increased efficiency.

(2) The standards must be designed for data distribution on Computer Disc-Read Only Memory (CD-ROM). This medium was selected for distribution because it is cost effective, efficient, and operates under established media standards.

(3) The standards must support interactive use of the media, so small computers can use the medium as "on-line" storage.

(4) The standards must be user oriented. They must support geographic modeling and analysis and graphics display. They must be useable by individuals oriented toward geographic information not computer science. The database access software will be available to users of data stored in the Vector Product Standards, and the source code will be available without restriction.

(5) The standards must be designed to work in a variety of computing environments regardless of manufacturer. They must work with a variety of vector data types regardless of the scale of the information or level of topological information.

(6) The standards must be compatible with Defense Mapping Agency's Digital Production System.

Prototyping. In order to insure that the standards would be user oriented and meet user requirements, an incremental, iterative prototyping approach was employed. With each prototype,

increasing quantities of data from several Operational Navigation Charts, hydrographic charts, and a topographic map were digitized and converted to a trial data structure and delivered to over 60 review sites. Review participants included the international development partners and a cross section of Defense Mapping Agency product users from the U.S. Department of Defense and Intelligence communities. A variety of product types were included in the prototypes to ensure that the standards apply to a range of geographic information systems applications at varying scales.

The first prototype, actually an ARC/INFO "mock-up" of the Digital Chart of the World, was delivered to the evaluation team in December 1989. This prototype consisted of a small section of Operational Navigation Chart G-18 (Santa Barbara, California) and a small patch of British Admiralty Chart 2693 (Harwich, England) The evaluators were provided with PC ARC/INFO software, an evaluation guide, hardcopy plots of the database, and complete instructions. By performing feature queries, viewing the data on the screen, and using a mock-up of the Digital Chart of the World software, the evaluators were able to assess the capabilities they wanted the application software to provide. Armed with this information the applications software programmers were able to quickly develop the software for Prototype 2, which was released in April 1989.

Prototype 2 used original "C" code and provided a newly implemented data structure. The data consisted of an expanded area of the two previous charts and also included National Ocean Service Charts 12245 and 12222 (Norfolk, Virginia) and a section of a 1:50,000 topographic map (Killeen, Texas). This prototype, which used 5 floppy disks to carry the data and one to store the software, provided evaluators with a good vehicle to analyze the database design and the new application software.

Prototype 3, released in early August, was the first prototype to use CD-ROM. The sources for prototype 3 were:

Operational Navigation Chart G-18 (California)
Operational Navigation Chart N-13 (Northern Australia)
Operational Navigation Chart E-18 (Ontario-Quebec, Canada)
BA 2( ,2 (Harwich, England)
National Ocean Service 12245 (Chesapeake Bay, U.S.)
Jet Navigation Chart 120 (Antarctica)

The application software was supplied on a separate floppy disk. This was the first prototype on the target medium and the first to implement the Vector Product Standards. The objectives of this prototype were two fold. First, to provide the standards, the software, and the data to the evaluation team; and secondly, to provide a test bed for the organization of the data, directories, and indices on CD-ROM. Along with the basic prototype sets Environmental Systems Research Institute added 8 test sets. These sets, with various structures of indices and directories, were

used to test data organization directly, with the CD-ROM in various drives. The sets provided calibration data for off-line tests to determine the optimum methods for storing this information. This will allow the Digital Chart of the World and other products to optimize the slow seek time and fast read time of the CD-ROM medium.

Recommendations from prototype 3 were then tested in prototype 4. Prototype 4 tested both the data structures and the production process. Using four Operational Navigation Charts from western Europe, E-01 (UK), F-01 (France, N. Spain), G-01 (S. Spain, N. Africa), and G-02 (N. Africa, S. Italy), prototype 4 tested the digitizing, feature tagging, transformation into the product format, indices and directories, CD-ROM production, and quality control throughout the process. This prototype and its specification constitute the first sample of the future Digital Chart of the World product.

This method of prototyping worked very well. Engineering ideas were quickly tested; the designs that didn't work well were quickly discovered and discarded, saving months of study and debate. Sound ideas were expanded and further proven with each iteration.

Studies.

A number of special studies have been initiated by Environmental Systems Research Institute to support prototype development. The Vector Product Standards has been separated into protocols, and Environmental Systems Research Institute has conducted research investigations into the various elements comprising each protocol. The lowest level protocols, those for CD-ROM, are following existing ISO 9660 standards and are not studied. The mid-level protocols consist of the spatial data structure, feature identification codes, and product content rules; at this level the studies include the Data Storage Structure Study, the Data Dictionary Study, the Digitizing Convention Study, the Aeronautical Study, and Elevation Data Study. The high level protocols consist of the data directory protocols; at this level the studies include the Tile Design Study, and the Indexing Studies: Coverage, Thematic, Gazetteer, and Spatial Query.

These studies were conducted in an iterative manner. Environmental Systems Research Institute initiated each study by investigating user needs, then conducted literature searches, consulted with area specialists, and issued a preliminary study which was sent to appropriate development participants for comment and enhancement. Responses were consolidated and used by Defense Mapping Agency to direct the contractor in refinement of the studies. The study findings were then folded into the prototypes for testing. The study delivery dates were staggered, each timed to mature coincident with the integration of the study element into the suite of Vector Product Standards. This method allowed

the designers access to the substantial knowledge and experience base of the cooperative partners and maximized their input to the final standard.

In this second year of development the special studies continue with investigations into Digital Chart of the World product maintenance issues and Digital Chart of the World product organization.

Vector Product Standards

Now that the Vector Product Standards development cycle is complete, the final suite has been submitted as U.S. Military Standards. While the Vector Product Standards will enforce commonality among future Defense Mapping Agency vector data sets, it has been designed for generic application. The Vector Product Standards consists of the following sub-sections:

Tiling. This standard addresses the units of area coverage into which a geographic database is divided. In large data sets it is awkward to work with the entire database as one unit. This is solved by dividing the data into manageable areas. The design of the tile standard is determined by performance criteria, data maintenance issues, data manageability, most acceptable access time, CD-ROM medium characteristics, and size and configuration of the computing environment.

Indices. Indices provide a mechanism for quickly finding information stored on the CD-ROM. The Vector Product Standards provides format and structure for Spatial Query Index, Tile Index, Coverage Index, Gazetteer Index, and Thematic Query Index.

Digitizing Conventions. The Vector Product Standards digitizing conventions define how to collect linear data as a continuous feature; how polygon features should be closed; and how intentionally coincident features should be stored in separate layers when required. The digitizing conventions do not define procedures for the physical transfer of data from hardcopy to softcopy.

Digital Marginalia. The Vector Product Standards defines how information, normally stored in the margin of a paper map, should be included in the database. This includes such information as data currency, compilation date, last revision date, and intended use.

Vector Product Format. The Vector Product Format is the heart of the Vector Product Standards. This subsection defines the conceptual and physical data model. In the Vector Product Format features are created from four basic primitives: nodes, edges, faces, and text. These are physically implemented in binary primitive and feature tables. There are four types of

the conceptual and physical data model. In the Vector Product Format features are created from four basic primitives: nodes, edges, faces, and text. These are physically implemented in binary primitive and feature tables. There are four types of primitive tables: Node Tables, Edge Tables, Face Tables, and Text Tables. These tables contain topologic relationships and coordinate data. There are three types of feature tables: Point Feature Tables, Line Feature Tables, and Area Feature Tables. These tables contain a primitive record ID and attribute values.

The Vector Product Standards is designed to support varying levels of topology and varying degrees of integration. This relational approach will allow data to be separated into layers. When applications do not require querying of the relationships among data types, data can be stored in separate layers, allowing efficient storage and access. When such querying is required, the layers can be integrated, carrying full topological relationships in the process. After much consideration the geo-relational model was selected because 1) a majority of the reviewers of previous prototype products expressed a desire for thematic separate data, and many future products may be very thematic in nature; 2) the natural indexing of relational tables will allow data producers to distribute data without building special purpose indices to speed access for each user's separate needs; 3) the minimum target configuration is best supported by simple file structures; 4) special purpose weapon systems can more easily use the data.

Once the Vector Product Standards becomes a U.S. Military Standard all U.S. Department of Defense vector product users will be required to conform to the standard. Command, Service, and Agency contracts for map digitization will require adherence to this standard. This will create a pool of digital mapping and charting data generated to a known standard, allowing data exchange within the U.S. Department of Defense and Intelligence communities and enhancing a common goal of interoperability.

U.S. Military Specification.

Where the U.S. Military Specification is designed to be as generic as possible, a product's U.S. Military Specification is quite constrained. The U.S. Military Specification defines the Data Dictionary, the features and attributes included in the product, the feature coding schemes employed, the convention for portrayal of special feature relationships such as complex features, and all other product specific requirements. The Digital Chart of the World U.S. Military Specifications references the Vector Product Standards and defines the specific implementations of this standard in Digital Chart of the World production. It specifies the level of topology, integration, data collection criteria, and a specific sub-set of the tiling and indexing within the standard. It also implements other standards such as U.S. Military Standards for accuracy and ISO 9660 for CD-ROM.

## THE DIGITAL CHART OF THE WORLD PRODUCT

### Background

The secondary goal of this development effort is no less ambitious than the primary goal of standards development; this is the development of the Digital Chart of the World product itself. Once completed, the data base will contain 1.5 - 2.0 gigabytes of data stored on 8 - 10 CD-ROMS, providing a comprehensive 1:1,000,000 scale digital base map of the world. The Digital Chart of the World will be the largest spatial database designed specifically for exploitation on relatively inexpensive PC class computers. Special indices will be provided to guide the user through the database.

Several factors led to the decision to produce a product with global coverage. First, there has long been interest in the International mapping community for a global database. The International Cartographic Association and its Working Group developed the concept for a World Digital Database for the Environmental Sciences in 1984 (International Cartographic Association, 1989); and the International Geographical Union and others have documented the requirements for a global database over the years. Secondly, a global database fills the digital data needs of a wide variety of users. Anyone with regional, strategic, or global interests will have use for this type of data; it is applicable to the operational problem sets addressed by Intelligence agencies, Command, Control, Communication, and Intelligence groups, the mission planning community, theater and strategic commands, and environmental and global monitoring organizations.

Finally, development of a global database provides the opportunity to comprehensively address the issues of tiling and indexing by uniquely challenging design efforts in these areas. Tiling must not only divide the global data set into manageable units, but these units must be nested, because different users work with different amounts of area coverage. Likewise, users of large databases encounter difficulties in determining which data they need to extract, and how to extract it. Elaborate indices, directories, and application software must be provided to guide users through the data extraction session. Large databases covering smaller regional areas also face these same problems. The Digital Chart of the World is providing the means to test and prove out global design solutions that can be applied to regional digital data sets in the future.

### Source

The 270 sheets in the 1:1,000,000 scale Operational Navigation Chart series comprise the primary source for the Digital Chart of the World. This series was selected because it is the largest

the World. This series was selected because it is the largest scale series that provides worldwide coverage; its scale allows expedient conversion to digital form; and Operational Navigation Charts are a product included in the Digital Production System. The series is also free from the restrictions governing the distribution of other products, allowing widespread dissemination of products derived from it. This primary source has been augmented by six charts from the 1:2,000,000 scale Jet Navigation Chart series in order to include Antarctica in the final product, making it truly a Digital Chart of the World. Conversion of hardcopy materials to digital form will provide the initial product; future editions and other Vector Product Standards compliant databases can be supplied by the Digital Production System.

The Operational Navigation Chart series provides standard topographic information consistent with basic 1:1,000,000 scale map products. This includes terrain elevation contours at 1,000 foot intervals; major transportation routes, including roads, single and multi-track railroads, and canals; populated places with city outlines; major drains and coastlines; and international boundaries

Approach

To create the digital database the chart production separates from the Operational Navigation Chart and Jet Navigation Chart series will be scanned and converted to vector form. The majority of separates (over 2000) will be scanned using a Scan-Graphics CF 1000 scanner. Three vectorization technologies are being employed to convert the raster data: the Optigraphics 3100 conversion system, Scan-Graphics Rave, and Hitachi CADCORE. Some separates will be hand digitized for efficiency. The vector data from these processors are then fed into the ARC/INFO system for creation of topological data structure, cartographic processing and editing, and quality control functions. Once the data have been converted into the Digital Chart of the World structure, the data will be transferred to the CD-ROM medium for distribution.

## THE DIGITAL CHART OF THE WORLD

Database.

The database will be designed to support both geographic analysis functions and graphics display. Although the data are being collected from cartographic source, it will be treated as if it were geographic information. Linear features such as roads and railroads that were deleted in the original source due to cartographic symbology, precedence, or displacement protocols will be connected. Additional source material is being used to connect transportation routes through built-up areas. Contours will be stored as continuous lines, closed at the ends and connected

on the hypsographic data, and many other spatial operations. The database, used in conjunction with a symbol library and a large display device, will provide briefing and planning graphics. Each CD-ROM will be richly indexed to allow comparatively rapid access. The Digital Chart of the World will be available for distribution by early 1992.

Software

The Digital Chart of the World project is providing a comprehensive database for commercial Geographic Information Systems, as well as an application software package. The application software, along with its source code, allow simple database query. It is a user friendly interface employing a pop-down windows menu that will provide immediate access to the data.

The software provides a variety of methods for defining a study area. The user can retrieve data by delimiting an area on a coverage map of the world; by entering a latitude and longitude position; or by selecting a feature name from gazetteer query. Once an area has been defined the operator can then select the feature types to be extracted from the database. A query list will be generated to allow all features of interest to be extracted at one time; the menu will display only those feature types available in the area of interest. This allows the information to be pulled from the database in one pass through the CD-ROM, speeding access time. Features can be added to the query set incrementally; however, this would require many re-reads of the disk and would slow access considerably.

Once feature selection is complete the information can be displayed to the terminal screen. The operator will have the option to change the color and size of the symbols, to highlight select features, and to zoom using set increments. Once displayed, the software will provide cursor position read-out in latitude-longitude and simple distance measuring capabilities. The software also provides the capability to save selected data from the database, the query selection set, and a bitmap of the graphics display on hard disk. The bitmap will allow quick recall and re-display of information previously extracted from the CD-ROM. The user may also off-load a simple plot file.

Although the application software provided is not a geographic information systems, it allows the user to examine the database and provides some basic utilities. The accompanying source code is provided to allow geographic information systems developers and other computer literate individuals to bring the Digital Chart of the World database into their operating environment.

## REFERENCES

Defense Mapping Agency, July 1990, Digital Chart of the World Interim Product Specification , MIL-D-89009.

Defense Mapping Agency, July 1990, Interim Digital Chart of the World Data Storage Structure Study.

Defense Mapping Agency, July 1990, Interim Digital Chart of the World Tiling Study.

Defense Mapping Agency, 1990, Interim Digital Chart of the World Indexing Study.

International Cartographic Association, May 1989, WDDES Strategy and Objectives.

Digital Geographic Information Working Group, October 1989, Digital Geographic Information Exchange Standards (DIGEST), Final Draft.

---

UNITED
NATIONS

E

Economic and Social Council

Distr.
LIMITED

E/CONF.83/L.28
21 January 1991

ENGLISH ONLY

TWELFTH UNITED NATIONS REGIONAL CARTOGRAPHIC
CONFERENCE FOR ASIA AND THE PACIFIC
Bangkok, 20-28 February 1991
Item 7 (d) of the provisional agenda*

# CARTOGRAPHIC DATA RETRIEVAL, ANALYSIS, DEPICTION, PRESENTATION AND PRODUCT GENERATION: GEOGRAPHICAL INFORMATION SYSTEMS

## How to benchmark your geographic information system

Paper submitted by the United States of America**

* E/CONF.83/1.

** Prepared by Stephen C. Guptill, United States Geological Survey.

91-01915 2059f (E)

/...

# HOW TO BENCHMARK YOUR GEOGRAPHIC INFORMATION SYSTEM

Geographic information systems (GIS) are emerging as the spatial data handling tools of choice for solving complex geographic problems. However, few guidelines exist to help potential users identify suitable hardware and software. Users must evaluate the merits of GIS technology according to standards and guidelines, software functions, hardware components, and benchmark testing, which identifies optimal hardware and software. By making users aware of all aspects of adopting GIS technology, they can decide if a GIS is an appropriate tool for them and which GIS should be used.

## HOW TO BENCHMARK YOUR GEOGRAPHIC INFORMATION SYSTEM

Some users feel that independent testing and rating of GIS, much like evaluations of personal-computer software by InfoWorld magazine, would aid their decisionmaking. Although independent testing may be useful for some types of evaluation, the user of GIS technology must be deeply involved in the evaluation process. Why is this the case? First, the process of benchmarking a GIS is a complex, difficult, and potentially expensive task. Second, and more important, GIS benchmarks should be driven by user requirements. In the end, users must benchmark their systems. This article will examine various components of GIS benchmarking, review what independent testers of GIS technology can provide, and discuss what users should do.

### User Requirements

GIS technology is successful when it meets users' needs comprehensively and consistently. Development of a successful GIS depends on well-defined user requirements; therefore, a user requirements analysis should be performed. The analysis should identify the operation, users, and data requirements of the existing system; identify potential uses and users of the GIS; define digital and hard-copy products required; evaluate work flow; estimate data base requirements that support GIS implementation; refine GIS product characteristics; calculate necessary production rates; and estimate data volumes. The analysis should also include a cost/benefit study.

The user requirements analysis documents how an organization functions and supports its users, and quantifies and defines an alternative operation on the basis of GIS technology. It does not address such issues as the organization's objectives, goals, or staff capabilities, which are not quantifiable technical issues. However, such factors could override conclusions based solely on technical considerations.

If, upon completing such an analysis and assessing the appropriateness of GIS for the organization's missions, it is determined that GIS technology should be used, evaluation criteria must be devised for selecting a specific GIS. The evaluation criteria should be incorporated in the specifications for benchmark testing.

Hands-on experience with GIS capabilities is valuable in developing evaluation criteria. One method for acquiring such experience is to perform a small-scale pilot project. The pilot project should be designed to test the ability of GIS technology to meet an organization's needs. Pilot tests are a source of realistic

data on production rates, memory and storage requirements, and user response to GIS products. Pilot test results can also be used to refine evaluation criteria that were derived from the user requirements analysis.

All or part of the GIS designed for the organization may be tested, depending on the organization's familiarity with GIS. Testing only portions of the GIS that are most critical to organizational needs, or that represent the elements with which the organization is least familiar, may be desirable.

## Identification of GIS Functions

Defining processing functions to meet specific needs is an important step in the design and evaluation of a GIS and is a direct outgrowth of the user requirements analysis. Identifying required functions often begins with a detailed list of GIS products and their specifications. Further analysis leads to identifying the type of processing functions needed to produce each product. My paper on evaluating GIS (1988, see references) includes a checklist of GIS software components.

Mandatory and desirable individual processing functions should be identified. Mandatory software capabilities, when merged with such specific application needs as required response time, accuracy, precision, product-generation frequency, and data volumes, lead directly to mandatory hardware capabilities.

Existing GIS's are extremely diverse in function and data base structure. Systems use various methods for digitizing, assigning, and storing attribute, coordinate, and topological information. The capability to manipulate, analyze, and display this information varies among systems. Capabilities of a given system are often oriented toward a specific function or application, such as computer-aided design and mapping, surveying, natural-resource management, terrain analysis, or image processing. Users must find the best match of system capabilities and user requirements. This is where benchmark testing can help.

Benchmarking is a process in which computer systems such as GIS are tested for function and performance. Benchmarking is accepted as part of the acquisition process by both private industry and the federal government. The success of benchmarking depends on the extent to which tests can be constructed that are representative of expected workloads. The following ten procedural steps for benchmark construction are described in Federal Information Processing Standard Publication 75, Guidelines on Constructing Benchmarks for ADP System Acquisition (U.S. National Bureau of Standards, 1980):

Step 1. Define benchmarking objectives and complete such preliminary activities as defining an agency's service, operational, and workload requirements.
Step 2. Quantify the present workload requirements.
Step 3. Obtain information from users on present applications and user forecasts of new or changing applications.
Step 4. Estimate future workload requirements.
Step 5. Categorize future workloads into distinct categories.
Step 6. Determine the relative contribution of each category.
Step 7. Scale each category. Weight the running times for each category's set of benchmarks according to its contribution.
Step 8. Represent workload categories with benchmarks. Select programs that represent the workload categories identified in step 5.
Step 9. Fine tune the mix of benchmarks on the present system.
Step 10. Prepare the benchmarking package, which consists of documentation of the mix of benchmarks and rules for the live test demonstration, and test the benchmarks.

A key point is that benchmarking must be done on products and capabilities identified through a user requirements analysis and must reflect estimated future data volumes. Users must also recognize that the preparation of a comprehensive benchmark may be a major task. Factors such as the size of the procurement must be weighed against the effort required to conduct benchmarking, and then a decision can be made on the appropriate level of testing.

## What to Test

The functions of GIS can be grouped into five categories: user interface, system/data base management, data base creation/data entry, data manipulation and analysis, and display and product generation. It is possible to develop function and performance requirements for each of these areas.

The user interface is the way operators communicate with the data base and GIS application modules. It consists of such software capabilities as menus, help screens, and graphic displays that simplify and organize the interaction between user and GIS software.

The data base management component provides the environment in which the GIS functions and the means by which the data are controlled. The system management environment is furnished by the operating system of the host computer. GIS data base functions parallel those of a non-spatial data base management system (DBMS), but with extensions beyond the addition, deletion,

revision, and Boolean retrieval capabilities of a standard DBMS. The GIS DBMS contains hardware and software facilities for storage, retrieval, and update of spatial information in alphanumeric and digital graphic forms, and incorporates storage structures to minimize data redundancy and to aid spatial searches. In addition, the GIS DBMS must have file management capabilities to handle a potentially large archive of geographic data files.

Data base creation/data entry refers to the process of bringing data into the GIS environment. A GIS data base contains a series of thematic categories or topics (layers) of information. These layers may contain information captured from aerial photographs, remote-sensing images, conventional maps, or other sources.

Data entry is the process of loading data into a GIS data base. Data in a computer-compatible format, such as digital remotely sensed data, can be loaded directly. A data base may also be created by digitizing or scanning maps or by digitizing information on aerial photographs to create a computer-readable data set. This effort is sometimes referred to as data capture.

Two types of data are generally collected: locational or geographic and accompanying feature attribute data. Locational information is usually digitized from existing graphics, maps, or images. Attributes identify what the features represent in the form of numeric or textual information such as soil type, feature name, or road class.

If spatial information capture and management are performed properly, the accuracy of the data will be maintained. The GIS user may then proceed to the primary GIS activities, data manipulation and analysis. Spatial analysis tools are used to model, make predictions, and reach conclusions about problems. Such analysis involves combining data from multiple spatial data categories and performing analytical, statistical, measurement, and other operations on the GIS data sets to transform the data into information suitable for a given application. Spatial analysis techniques include compositing areas, performing proximity searches, analyzing networks, and clustering and aggregation operations. These operations are ideally performed in an interactive mode on the spatial component of the data, as well as on non-spatial attribute data. Such operations can range in complexity from simple Boolean queries to reclassification and creation of new map displays.

The typical GIS has capabilities for display and product generation. Maps, charts, graphs, and tables can be produced, and displays can range in complexity from tabular reports and simple monochrome plots to publication-quality, three-dimensional color graphics.

GIS hardware components can be divided into six major categories of devices: data capture/input, processing, communications/networking, interactive display and edit, external storage, and hard-copy output. These categories can be further divided into hundreds of hardware components used in an information processing system. Hardware components can be tested against a number of measures: performance (speed of the central processing unit, rate of graphics display), accuracy (precision of a digitizing table, resolution of a plotter), and capability (maximum plot size, disk storage size). The total specification of all aspects of GIS hardware is laborious and probably not required. Users should focus on aspects of hardware specifications that directly affect performance.

## Testing

Once users decide what to test, they must consider how to perform the testing. Do they test a specific function such as polygon overlay with or without sliver removal, or do they test a number of system components simultaneously? For example, the Structured Query Language (SQL) statements that follow test a variety of system components including the ability to optimize queries, the effectiveness of indexes, and the performance of the spatial operator WITHIN. This series of statements asks the GIS to find all public schools located within areas of high radon concentration.

```
SELECT *
FROM BLDG
WHERE BLDG.LOC WITHIN RADON.POLY
    AND BLDG.LOC IN
        (SELECT BLDG.LOC
        FROM BLDG
        WHERE BLDG.TYPE = 'PUBLIC SCHOOL')
    AND RADON.POLY IN
        (SELECT RADON.POLY
        FROM RADON
        WHERE RADON.POT > 4.0)
```

Another factor to consider is accuracy. How do users know if the results are correct? One way is to test GIS functions with mathematically generated data sets for which correct answers are known. For example, two sets of concentric ellipses could be intersected as one test of the correctness of a polygon overlay operation.

Users must also remember that test results are dependent on the data sets used. "Stress testing" systems with large data sets may reveal problems. Systems that behave well with modest-size data sets may rebel when processing large data sets.

## The Bottom Line

The traditional methods of acquiring, storing, and analyzing spatially referenced data are becoming too costly and inflexible in meeting today's information requirements. Even so, potential users of GIS should understand all aspects before adopting this new technology. Is GIS an appropriate tool for the intended application, and if so, which GIS should be used?

Users and independent testers can work together to provide answers to this question. Users must ensure that the system has the software functions required to support their applications, and that their hardware configuration is adequate to support their software and provide a path for growth. Independent testers should ensure that key GIS functions produce accurate results, and should "stress test" key GIS functions by using a standardized set of test data.

Evaluating GIS technology will help managers and technical specialists understand GIS technology from a realistic perspective.

## References

Chrisman, N.R. 1987. "Efficient Digitizing through the Combination of Appropriate Hardware and Software for Error Detection and Editing." International Journal of Geographical Information Systems, vol. 1, no. 3, pp. 265-277.

Gohagan, J.K. 1980. Quantitative Analysis for Public Policy. New York: McGraw-Hill.

Goodchild, M.F., and B.R. Rizzo. 1987. "Performance Evaluation and Workload Estimation for Geographic Information Systems." International Journal of Geographical Information Systems, vol. 1, no. 1, pp. 67-76.

Guptill, S.C., ed. 1985. Proceedings of the Workshop on Methods and Techniques for Digitizing Data. U.S. Geological Survey Open-File Report 85-307.

Guptill, S.C., ed. 1988. A Process for Evaluating Geographic Information Systems. Federal Interagency Coordinating Committee on Digital Cartography, Technology Exchange Working Group, Technical Report 1, U.S. Geological Survey Open-File Report 88-105.

Smith, Terence, Sydhakar Menon, Jeffrey Star, and John Estes. 1987. "Requirements and Principles for the Implementation and Construction of

Large-Scale Geographic Information Systems." International Journal of Geographical Information Systems, vol. 1, no. 1, pp. 13-31.

Stefanovic, Pavao, and Jane Drummond. 1987. "Selection and Evaluation of Computer-Assisted Mapping and Geo Information Systems." ITC Journal, no. 1, pp. 39-44.

Tomlinson, R.F., and A.R. Boyle. 1981. "The State of Development of Systems for Handling Natural Resources Inventory Data." Cartographica, vol. 18, no. 4, pp. 65-95.

U.S. National Bureau of Standards. 1977. Guidelines for Benchmarking ADP Systems in the Competitive Procurement Environment. Federal Information Processing Standard 42 (FIPS PUB 42-1).

U.S. National Bureau of Standards. 1980. Guidelines on Constructing Benchmarks for ADP System Acquisition. Federal Information Processing Standards Publication 75 (FIPS-PUB-75).

---